

University of Groningen

## A captivating snapshot of standardized testing in early childhood

Frans, Niek

DOI:  
[10.33612/diss.95431744](https://doi.org/10.33612/diss.95431744)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Frans, N. (2019). *A captivating snapshot of standardized testing in early childhood: on the stability and utility of the Cito preschool/kindergarten tests*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen. <https://doi.org/10.33612/diss.95431744>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# References

- Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation*, 33(1), 50–68. <https://doi.org/10.1016/j.stueduc.2007.01.005>
- Allison, P. D. (2009). Missing Data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 72–90). California: Sage Publications, Inc.
- Asendorpf, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality*, 6(2), 103–117. <https://doi.org/10.1002/per.2410060204>
- Bagnato, S. J., Mclean, M., Macy, M., & Neisworth, J. T. (2011). Identifying instructional targets for early childhood via authentic assessment: Alignment of professional standards and practice-based evidence. *Journal of Early Intervention*, 33(4), 243–253. <https://doi.org/10.1177/1053815111427565>
- Barnes, N., Fives, H., & Dacey, C. (2015). Teachers' beliefs about assessment. In H. Fives & M. Gregoire Gill (Eds.), *International Handbook of Research on Teachers' Beliefs* (pp. 284–300). New York.
- Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116. <https://doi.org/10.1016/j.tate.2017.02.017>
- Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, 333(6045), 975–979. <https://doi.org/10.1126/science.1204534>
- Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, 1(4), 1–31. <https://doi.org/10.1177/2332858415616358>
- Black, P. J. (2002). *Testing, friend or foe?: The theory and practice of assessment and testing*. London: Routledge.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Formative and summative assessment : Can they serve learning together? In *Annual meeting of the American Educational Research Association*. Chicago.
- Bonner, S. M. (2016). Teachers' perceptions about assessment: competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (1st ed., pp. 21–39). New York: Routledge.
- Bordignon, C. M., & Lam, T. C. M. (2004). The early assessment conundrum: Lessons from the past, implications for the future. *Psychology in the Schools*, 41(7), 737–749. <https://doi.org/10.1002/pits.20019>
- Bornstein, M. H., Brown, E., & Slater, A. (1996). Patterns of stability and continuity in attention across early infancy. *Journal of Reproductive and Infant Psychology*, 14(3), 195–206. <https://doi.org/10.1080/02646839608404517>
- Bornstein, M. H., Hahn, C., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, 24(3), 267–304. <https://doi.org/10.1177/0142723704045681>

## References

- Bornstein, M. H., Hahn, C., Putnick, D. L., & Suwalsky, J. T. D. (2014). Stability of core language skill stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development, 85*(4), 1346–1356. <https://doi.org/10.1111/cdev.12192>
- Bornstein, M. H., & Lansford, J. E. (2013). Assessing Early Childhood Development. In P. Rebello Britto, P. L. Engle, & C. M. Super (Eds.), *Handbook of Early Childhood Development Research and Its Impact on Global Policy* (pp. 353–361). Oxford Scholarship Online. <https://doi.org/10.1093/acprof:oso/9780199922994.001.0001>
- Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology, 48*(2), 477–491. <https://doi.org/10.1037/a0025889>
- Bracken, B. A., & Walker, K. C. (1997). The utility of intelligence tests for preschool children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (1st ed., pp. 484–502). New York: The Guilford Press.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*(1), 166–170. <https://doi.org/10.2466/pr0.99.1.166-170>
- Brown, G. T. L. (2008). Teachers' conceptions of assessment. In Gavin T. L. Brown (Ed.), *Conceptions of Assessment: Understanding What Assessment Means to Teachers and Students*. (pp. 91–118). New York: Nova.
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation, 6*(12), 68–91.
- Brown, G. T. L., & Hattie, J. (2012). The benefits of regular standardized assessment in childhood education: Guiding improved instruction and learning. *Contemporary Debates in Childhood Education and Development, (January)*, 287–292. <https://doi.org/10.4324/9780203115558>
- Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5–6), 307–320. <https://doi.org/10.1016/j.ijer.2011.10.003>
- Brown, G. T. L., Lake, R., & Matters, G. (2009). Assessment policy and practice effects on New Zealand and Queensland teachers' conceptions of teaching. *Journal of Education for Teaching, 35*(1), 61–75. <https://doi.org/10.1080/02607470802587152>
- Burger, K. (2010). How does early childhood care and education affect cognitive development ? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly, 25*, 140–165. <https://doi.org/10.1016/j.ecresq.2009.11.001>
- Centraal Bureau voor de Statistiek. (2012). *Jaarboek onderwijs in cijfers 2012*. Den Haag. Retrieved from <http://www.cbs.nl/NR/rdonlyres/3036B4E1-A671-4C9E-95BF-90C0493B4CD9/0/2012f162pub.pdf>
- Colpin, M., Gysen, S., Jaspaert, K., Heymans, R., Van den Branden, K., & Verhelst, M. (2006). *Studie*

- naar de wenselijkheid en haalbaarheid van de invoering van centrale taaltoetsen in Vlaanderen in functie van gelijke onderwijskansen.* Leuven.
- COTAN. (2010). *Toelichting bij de beoordeling Rekenen-Wiskunde Groep 3 t/m 8 LOVS Cito.* Amsterdam: COTAN.
- COTAN. (2011). *Toelichting bij de beoordeling Rekenen voor Kleuter Groep 1 en 2 LOVS Cito.* Amsterdam.
- COTAN. (2013). *Toelichting bij de beoordeling Taal voor Kleuters (TvK).* Utrecht.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Daniels, L. M., Poth, C., Papile, C., & Hutchison, M. (2014). Validating the Conceptions of Assessment-III scale in Canadian preservice teachers. *Educational Assessment*, 19(2), 139–158. <https://doi.org/10.1080/10627197.2014.903654>
- Darrah, J., & Hodge, M. (2003). Stability of serial assessments of motor and communication abilities in typically developing infants—implications for screening. *Early Human Development*, 72(2), 97–110. [https://doi.org/10.1016/S0378-3782\(03\)00027-6](https://doi.org/10.1016/S0378-3782(03)00027-6)
- De Wijs, A., Kamphuis, F., Kleintjes, F., & Tomesen, M. (2010). *Wetenschappelijke verantwoording Spelling voor groep 3 tot en met 6.* Arnhem.
- DeLuca, C., & Hughes, S. (2014). Assessment in early primary education: An empirical study of five school contexts. *Journal of Research in Childhood Education*, 28(4), 441–460. <https://doi.org/10.1080/02568543.2014.944722>
- Den Elt, M. E., Van Kuyk, J. J., & Meijnen, G. W. (1996). Culture and the kindergarten curriculum in the Netherlands. *Early Child Development and Care*, 123, 15–30.
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health*, 20(2), 116–125. <https://doi.org/10.1111/camh.12072>
- Dollaghan, C. A., & Campbell, T. F. (2009). How well do poor language scores at ages 3 and 4 predict poor language scores at age 6? *International Journal of Speech-Language Pathology*, 11(5), 358–365. <https://doi.org/10.1080/17549500903030824>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- DUO. (2013). Toelichting Gewichtenregeling basisonderwijs. Retrieved from [https://www.duo.nl/Images/Toelichting\\_gewichtenregeling\\_basisonderwijs\\_29\\_april\\_2013\\_tcm7-39943.pdf](https://www.duo.nl/Images/Toelichting_gewichtenregeling_basisonderwijs_29_april_2013_tcm7-39943.pdf)
- DUO. (2017). *Leerlingen po zoals geregistreerd in BRON.*
- DUO. (2018). *Adressen van alle schoolvestigingen in het basisonderwijs.*
- Dutch Eurydice Unit. (2007). *The education system in the Netherlands 2007.* The Hague.

## References

- Einarsdóttir, J. T., Björnsdóttir, A., & Símonardóttir, I. (2016). The predictive value of preschool language assessments on academic achievement: A 10-year longitudinal study of Icelandic children. *American Journal of Speech-Language Pathology*, 25, 67–79.
- European Commission/EACEA/Eurydice. (2015). *Early childhood education and care systems in Europe. National information sheets – 2014/15*. Brussels. <https://doi.org/10.2797/48986>
- Faber, M., Van Geel, M., & Visscher, A. (2013). *Digitale leerlingvolgsystemen als basis voor opbrengstgericht werken in het primair onderwijs*. Enschede.
- Feenstra, H., Kamphuis, F., Kleintjes, F., & Krom, R. (2010). *Wetenschappelijke verantwoording Begrijpend lezen voor groep 3 tot en met 6*. Arnhem.
- Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2014). Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology*, 106(2), 482–498. <https://doi.org/10.1037/a0034444>
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40, 1–37. Retrieved from <http://www.jstatsoft.org/v40/i04>
- Geary, D. C. (2006). Development of Mathematical Understanding. In D. Kuhn, R. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology, Cognition, Perception, and Language* (6th ed., pp. 777–810). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Gelderblom, G., Schildkamp, K., Pieters, J., & Ehren, M. (2016). Data-based decision making for instructional improvement in primary education. *International Journal of Educational Research*, 80, 1–14. <https://doi.org/10.1016/j.ijer.2016.07.004>
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, 38(4), 293–304.
- Gilliam, W. S., & Frede, E. (2012). Accountability and program evaluation in early education. In R. C. Pianta, W. Steven Barnett, L. M. Justice, & S. M. Sheridan (Eds.), *Handbook of Early Childhood Education* (pp. 77–91). New York: The Guilford Press.
- Goorhuis, S. M., & Schaerlaekens, A. M. (2000). *Handboek Taalontwikkeling, Taalpathologie en Taaltherapie bij Nederlandssprekende Kinderen* (2nd ed.). Leusden: De Tijdstroom.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Graham, J. W. (2012). *Missing Data: Analyses and Design*. New York: Springer. <https://doi.org/10.1007/978-1-4614-4018-5>
- Guralnick, M. (2005). Early intervention for children with intellectual disabilities: Current knowledge and future prospects. *Journal of Applied Research in Intellectual Disabilities*, 18, 313–324. <https://doi.org/10.1111/j.1468-3148.2005.00270.x>
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: tensions between the needs of schools and students. *Assessment in Education: Principles, Policy*

- & *Practice*, 16(3), 365–381. <https://doi.org/10.1080/09695940903319745>
- Hartmann, D. P., Pelzel, K. E., & Abbott, C. B. (2011). Design, measurement, and analysis in developmental research. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science, an advanced textbook* (6th ed., pp. 109–198). New York: Psychology Press.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3–56. <https://doi.org/10.1006/reec.1999.0225>
- Janssen, J., Verhelst, N. D., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8*. Arnhem.
- Kagan, J. (1971). *Change and continuity in infancy*. New York: John Wiley & Sons, Inc.
- Kagan, J. (1980). Perspectives on continuity. In G. B. Orville & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). London: Harvard University Press.
- Keuning, J., Hilte, M., & Weekers, A. (2014). Begrijpend leesprestaties onderzocht - Een analyse op basis van Cito dataretour. *Tijdschrift Voor Orthopedagogiek*, 53, 2–13.
- Keuning, J., Van Boxtel, H., Lansink, N., Visser, J., Weekers, A., & Engelen, R. (2015). *Actualiteit en kwaliteit van normen. Een werkwijze voor het normeren van een leerlingvolgsysteem*. Arnhem.
- Kilderry, A. (2015). The intensification of performativity in early childhood education. *Journal of Curriculum Studies*, 47(5), 634–653. <https://doi.org/10.1080/00220272.2015.1052850>
- Kim, J., & Suen, H. K. (2003). Predicting children's academic achievement from early assessment scores: A validity generalization study. *Early Childhood Research Quarterly*, 18(4), 547–566. <https://doi.org/10.1016/j.ecresq.2003.09.011>
- Koerhuis, I. (2010). *Handleiding rekenen voor kleuters groep 1 en 2 [Manual mathematics for kindergartners]*. Arnhem: Cito.
- Koerhuis, I., & Keuning, J. (2011). *Wetenschappelijke verantwoording van de toetsen Rekenen voor kleuters*. Arnhem: Cito.
- La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70(4), 443–484. <https://doi.org/10.3102/00346543070004443>
- Lam, R. (2013). Formative use of summative tests: using test preparation to promote performance and self-regulation. *The Asia-Pacific Educational Researcher*, 22(1), 69–78. <https://doi.org/10.1007/s40299-012-0026-0>
- Lansink, N. (2009). *Handleiding taal voor kleuters [Manual language for kindergartners]*. Arnhem: Cito.
- Lansink, N., & Hemker, B. T. (2012). *Wetenschappelijke Verantwoording van de toetsen Taal voor kleuters voor groep 1 en 2 uit het Cito Volgsysteem primair onderwijs*. Arnhem: Cito. Retrieved from <http://toetswijzer.kennisnet.nl/html/tg/18.pdf>
- Law, J., Boyle, J., Harris, F., Harkness, a, & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Developmental Medicine and Child Neurology*, 42(3), 190–200. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10755459>



## References

- Lerner, R. M., Lewin-Bizan, S., & Alberts Warren, A. E. (2011). Concepts and theories of human development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science, an advanced textbook* (6th ed., pp. 3–50). New York: Psychology Press.
- Leseman, P. (2004). De toegevoegde waarde van vroeg testen. *Pedagogiek*, 24(1), 3–11.
- Mashburn, A. J., & Henry, G. T. (2004). Assessing school readiness: validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice*, 23(4), 16–30. <https://doi.org/10.1111/j.1745-3992.2004.tb00165.x>
- McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*, 52(1), 1–12. <https://doi.org/10.2307/1129210>
- Meisels, S. J., Steele, D., & Quinn, K. (1989). *Testing, tracking, and retaining young children: An analysis of research and social policy*.
- Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic assessment for infants and toddlers: Exploring the reliability and validity of the ounce scale. *APPLIED DEVELOPMENTAL SCIENCE*, 14(2), 55–71. <https://doi.org/10.1080/10888691003697911>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
- Mroczek, D. K. (2007). The analysis of longitudinal data in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 543–556). New York: The Guilford Press.
- Nagle, R. J. (2000). Issues in preschool assessment. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (3rd ed., pp. 19–32). Boston: Allyn & Bacon.
- Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(4), 262–279.
- Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics*, 117(2), e298–319. <https://doi.org/10.1542/peds.2005-1467>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. <https://doi.org/10.1080/09695940701478321>
- Newton, P. E., & Shaw, S. D. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, 23(2), 178–197. <https://doi.org/10.1080/0969594X.2015.1037241>
- OECD. (2008). *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools* (Vol. 2008). OECD Publishing.
- Oosterhoff, A., Minnaert, A. E. M. G., Oenema-Mostert, C. E., & Goorhuis-Brouwer, S. (2014). *Constrained or sustained by demands? Personal feelings of professional autonomy in early childhood education. Poster session presented at the 24th EECERA conference*. Crete, Greece.
- Papenburg, I. (2015). Kleutertoetsen, wie helpt me met mijn hoofdbreken? Retrieved October 11, 2018, from <https://citoblog.wordpress.com/2015/01/27/kleutertoetsen-wie-helpt-me-met-mijn-hoofdbreken/>

- Pellegrino, L. (2012). Patterns in development and disability. In M. L. Batshaw, N. J. Roizen, & G. R. Lotrecchiano (Eds.), *Children with disabilities* (7th ed., pp. 231–242). Baltimore: Brookes Publishing Co.
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: the Spanish instance, building upon Black and Wiliam (2005). *Curriculum Journal*, 18(1), 27–38. <https://doi.org/10.1080/09585170701292133>
- Rijksoverheid. (2018). Wat is er precies besloten over de kleutertoets ? Retrieved February 6, 2019, from <https://abonneren.rijksoverheid.nl/nieuwsbrieven/archief/artikel/33/6114fe75-dc99-4354-92ee-0c1a98a6d8f9/1fd4b772-539f-41df-9ce3-2486d1947a8e>
- Roberts-Holmes, G., & Bradbury, A. (2016). The datafication of early years education and its impact upon pedagogy. *Improving Schools*, 19(2), 119–128. <https://doi.org/10.1177/1365480216651519>
- Rog, M. R. J., Bisschop, R., van Dijk, J., Voordewind, J. S., & Klaver, J. F. (2013). *Motie van het lid Rog c.s. over een landelijk genormeerde kleutertoets*. 's Gravenhage. Retrieved from [http://www.tweedekamer.nl/downloads/document/index.jsp?id=cec73f89-926a-4c86-824b-956dfd8695c6&title=Motie van het lid Rog c.s. over een landelijk genormeerde kleutertoets .pdf](http://www.tweedekamer.nl/downloads/document/index.jsp?id=cec73f89-926a-4c86-824b-956dfd8695c6&title=Motie%20van%20het%20lid%20Rog%20c.s.%20over%20een%20landelijk%20genormeerde%20kleutertoets.pdf)
- Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: replication and extension using a nationwide Canadian survey. *Developmental Psychology*, 46(5), 995–1007. <https://doi.org/10.1037/a0018880>
- Rovict B.V. (n.d.). Leerlingen, BRON. Retrieved from [http://www.rovict.nl/downloads/FAQ\\_NNCA.pdf](http://www.rovict.nl/downloads/FAQ_NNCA.pdf)
- RTL. (2017). *Eindtoets cijfers*.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.
- Rudinger, G., & Rietz, C. (1998). The neglected time dimension? Introducing a longitudinal model testing latent growth curves, stability, and reliability as time bound processes. *Methods of Psychological Research Online*, 3(2), 109–130.
- Scarborough, H. S. (2009). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In F. Fletcher-Campbell, J. Soler, & G. Reid (Eds.), *Approaching difficulties in literacy development: assessment, pedagogy and programmes* (pp. 23–38). London: Sage.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, 37, 49–54. <https://doi.org/10.1016/j.stueduc.2011.03.008>
- Serafini, F. (2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher*, 54(4), 384–393.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 15–22.
- Shepard, L. A. (1994). The challenges of assessing young children appropriately. *The Phi Delta Kappan*, 76(3), 206–212. Retrieved from <http://www.jstor.org/stable/20405297>



## References

- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5–24. <https://doi.org/10.1111/j.1745-3992.1997.tb00585.x>
- Shepard, L. A., Kagan, S. L., Wurtz, E., Graves, W., Patton, P. E., Romer, R., ... Whitman, C. T. (1998). *Principles and recommendations for early childhood assessments*. Darby: DIANE Publishing.
- Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy of Sciences - National Research Council. Retrieved from <http://eric.ed.gov/?id=ED446866>
- Siegler, R. S. (2002). Variability and infant development. *Infant Behavior and Development*, 25(4), 550–557. [https://doi.org/10.1016/S0163-6383\(02\)00150-9](https://doi.org/10.1016/S0163-6383(02)00150-9)
- Smooenburg, B. Y. (2013). *Zie je ze GROElen!? Hoe observatie van authentiek gedrag , evaluatie en planning met GOLD-NL het pedagogisch handelen van voorschoolse professionals kan versterken*. Groningen: Rijksuniversiteit Groningen.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage Publications, Inc.
- Snow, K. L. (2006). Measuring school readiness : Conceptual and practical considerations. *Early Education and Development*, 17(1), 7–41. <https://doi.org/10.1207/s15566935eed1701>
- Steenbeek, H., & Van Geert, P. (2018). Assessing young children ' s learning and behavior in the classroom : A complexity approach. In M. Fleeer & B. Van Oers (Eds.), *International Handbook of Early Childhood Education* (pp. 1279–1299). Berlin: Springer. <https://doi.org/10.1007/978-94-024-0927-7>
- Sterba, S. K., & Pek, J. (2012). Individual influence on model selection. *Psychological Methods*, 17(4), 582–599. <https://doi.org/10.1037/a0029253>
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466–478. <https://doi.org/10.1111/j.1467-8527.2005.00307.x>
- Tiekstra, M., Bergwerff, L., & Minnaert, A. (2017). Voices from practice: When is the gap between diagnosis and intervention apparent? *Educational & Child Psychology*, 34(1), 55–66.
- Tisak, J., & Meredith, W. (1990). Descriptive and associative developmental models. In A. Von Eye (Ed.), *Statistical methods in longitudinal research 2* (pp. 387–406). Academic Press. Retrieved from <http://ebookcentral.proquest.com>
- Torrance, H. (1997). Assessment, Accountability, and Standards: Using Assessment to Control the Reforming of Schooling. In A. H. Halsey, H. Lauder, P. Brown, & A. Stuart Wells (Eds.), *Education, Culture, Economy, and Society* (pp. 320–337). Oxford.
- Tsagris, M., Beneki, C., & Hassani, H. (2014). On the folded normal distribution. *Mathematics*, 12–28. <https://doi.org/10.3390/math2010012>
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE : Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3).
- Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, 20(11), 1–19. Retrieved from <http://www.jstatsoft.org/v20/i11/>

- Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, 48(5), 1–27. Retrieved from <http://www.jstatsoft.org/v48/i05/>
- Van Eerde, H. A. A. (2009). Rekenen-wiskunde en taal: een didactisch duo. *Reken-Wiskundeonderwijs: Onderzoek, Ontwikkeling, Praktijk*, 28(3), 19–32.
- Van Engelshoven, I. (2018). *Uitvoering regeerakkoord t.a.v. kleutertoetsen*. Den Haag.
- Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2014). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0086817>
- Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *One-Parameter Logistic Model OPLM*. Arnhem.
- Verhelst, N. D., Verstralen, H. H. F. M., & Eggen, T. J. H. M. (1991). *Finding starting values for the item parameters and suitable discrimination indices in the One-Parameter Logistic Model*. Arnhem.
- Vincent-Lancrin, S. (2010). OECD mapping of longitudinal information systems: preliminary results. In *Educational information systems for innovation and improvement*. New York: OECD, Centre for Educational Research and Innovation.
- Vlug, K. F. M. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies*, 2, 287–306.
- Werkgroep en Steungroep Kleuteronderwijs. (2013). *Zwartboek, kleuters in de knel!*
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399. <https://doi.org/10.1002/sim.4067>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr00799>
- Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.
- Wohlwill, J. F. (1980). Cognitive development in childhood. In G. B. Orville & J. Kagan (Eds.), *Constancy and change in human development* (pp. 359–444). London: Harvard University Press.
- Zubrick, S. R., Taylor, C. L., & Christensen, D. (2015). Patterns and predictors of language and literacy abilities 4–10 years in the longitudinal study of Australian children. *PLoS ONE*, 10(9), 1–29. <https://doi.org/10.1371/journal.pone.0135612>

